Nano Communication Networks 6 (2015) 133–144

Contents lists available at ScienceDirect

Nano Communication Networks

journal homepage: www.elsevier.com/locate/nanocomnet

Contribution of canonical feed-forward loop motifs on the fault-tolerance and information transport efficiency of transcriptional regulatory networks

Ahmed F. Abdelzaher [a], Michael L. Mayo [b], Edward J. Perkins [b], Preetam Ghosh [a,*]

[a] Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA

[b] Environmental Laboratory, US Army Engineer Research and Development Center, Vicksburg, MS, USA

ARTICLE INFO

Article history:
Received 16 March 2015
Accepted 1 April 2015
Available online 21 April 2015

Keywords:
Shortest path
Centrality
Graph randomization
Clustering coefficient
Feed-forward loop

ABSTRACT

Motifs and degree distribution in transcriptional regulatory networks play an important role towards their fault-tolerance and efficient information transport. In this paper, we designed an innovative in silico canonical feed-forward loop motif knockout experiment in the transcriptional regulatory network of *E. coli* to assess their impact on the following five topological features: average shortest path, diameter, closeness centrality, global and local clustering coefficients. Additional experiments were conducted to assess the effects of such motif abundance on *E. coli*'s resilience to nodal failures and the end-to-end transmission delay. The purpose of this study is two-fold: (i) motivate the design of more accurate transcriptional network growing algorithms that can produce similar degree and motif distributions as observed in real biological networks and (ii) design more efficient bio-inspired wireless sensor network topologies that can inherit the robust information transport properties of biological networks.

Specifically, we observed that canonical feed forward loops demonstrate a strong negative correlation with the average shortest path, diameter and closeness centralities while they show a strong positive correlation with the average local clustering coefficient. Moreover, we observed that such motifs seem to be evenly distributed in the transcriptional regulatory network; however, the direct edges of multiple such motifs seem to be stitched together to facilitate shortest path based routing in such networks.

1. Introduction

The ability for networks to rewire its links was introduced by biologists when they realized that biological networks can resist external perturbations, yet proceed with their natural activities. This property was referred to as 'Biological Robustness' [1], and was mainly attributed to these network's topological features [2]. For example in the proposed bio-inspired self organizing wireless sensor and actuator network [3], edge rewiring must guarantee optimal topological preservation in case of nodal collapse.

Among the features that aid in network dynamics, is the fact that biological networks are sparse [4], or loosely connected. In such networks, degree distributions can be expressed using a power law, $p(k) \sim k^{-\gamma}$ [5], meaning, a steep negative slope will result from the bi-log plot of the different nodal degrees vs. the nodes that have such degrees. Consequently, few nodes have degrees much higher

* Corresponding author.
E-mail addresses: abdelzaheraf@vcu.edu (A.F. Abdelzaher), Michael.L.Mayo@usace.army.mil (M.L. Mayo), Edward.J.Perkins@usace.army.mil (E.J. Perkins), pghosh@vcu.edu (P. Ghosh).

http://dx.doi.org/10.1016/j.nancom.2015.04.002

than the average degree (i.e. hubs) [6], while the bulk of the remaining nodes have degrees much lower than the average, resulting in loosely connected components. For example, most gene nodes in the transcriptional network of Escherichia coli (herein E. coli), have single incoming links and no outgoing ones.

Most biological networks have $2 < \gamma < 3$ [5]—a property which classifies a network as 'scale-free' [6]. The significance of networks having hub to low degree orientations, can be expressed through the reduced probability of having a detrimental random attack. Most random attacks will result in loss of minimal links [7]. Similarly with edge rewiring, the relative overall damage to the entire network will not be too high. However intentional hub attacks can be very costly and can result in disconnecting the network [8].

Other network classifications fall under two other major categories, namely random (ER) [9] and small-world (SW) [10] networks. The preceding considers links connecting two nodes at equal probabilities during growth, while the latter is inclined towards minimizing the number of hops between pairs of nodes. In contrast with scale-free networks, systems grow using 'the rich get richer and the poor get poorer' phenomena. ER and SW networks typically have tightly connected components, which are different from that of scale-free networks. Nodal degrees are almost equal to the average degree, therefore making random attacks as equally costly as intentional attacks [6].

Another important aspect that aids in biological robustness, is the existence of 3–6 nodal substructures known as 'motifs' [11]. For example, ecologists believe that synthetic communities forming motifs can be inserted with weak interactions to increase the community's stability [12]. These substructures are labeled 'significant' because their abundance in the real networks is much higher as compared with their numbers in multiple randomized versions of these networks [11,13]. One of the major goals of this paper is to demonstrate that motif abundance aids in the overall information transport in such transcriptional regulatory networks (herein TRNs).

Understanding the underlying architecture of TRNs can provide insights in disease dynamics and drug development [14,15]. In a TRN, nodes portray the genes in a cell, and a set of directed links that correspond to interacting pairs or genes [16]. Interactions could either represent translation or transcription [17]. Unlike engineered networks, TRNs exhibit biological robustness [1,2] due to their tolerance of noise during gene expression [18]. This phenomena arises from feed-back control nodal arrangements and repetitive substructures [18], or motifs.

Among the significant 3-node motifs in TRNs, the feed-forward loop (or FFL) has received the most attention. An FFL consists of a transcription factor A that regulates another transcription factor B and a gene C, while B co-regulates C. This topology allows it to deliver essential tasks like generating pulses, irreversible speed ups and signal delays [19].

At the nodal level, research has been directed towards correlating robustness with disturbances in the network. For example, the effect of nodal collapse has been modeled to show a decrease in network efficiency [20,21]—which is defined as the inverse of the average shortest path in the network [22]. The effects of successive random node deletions on the network diameter were studied for scale-free and ER networks [6,7]. Others model disturbances in the form of connection failures [23], or partial inactivation [24], where the length of the links in the TRN of E. coli were increased to resemble interaction delays.

At the motif level, much attention has been focused towards understanding motif functions. Different motif configurations have been investigated using mathematical models of transcription and translation to understand the relationship between coupling and function of embedded motifs [25,26]. Experiments with E. coli have been conducted where transcription factors were rewired and the tolerance of the bacteria was analyzed [27].

However, little is known regarding the *motif distribution* in the network and its contributions towards robustness, particularly in terms of information transport. In this paper, we address this issue by designing *in-silico* FFL knock-outs in E. coli, while preserving the individual in-, out- and cumulative degrees of the nodes. After every successive FFL deletion, different topological features of the resulting TRNs were recorded. Then, Pearson's correlation coefficient is used to determine the correlation between each pair of metrics studied here.

2. Motivation

2.1. Bio-inspired wireless sensor networks

Wireless sensor networks form a special class of engineered systems wherein sensor nodes forward data packets that are routed through adjacent sensors to a sink capable of processing the sensed information. Resemblance between gene regulation systems and wireless sensor networks (herein WSNs) can be described through transcription, where genes process signals from adjacent neighbors in the form of transcription factors that excite/repress other genes by generating mRNA molecules. Nodes in a TRN interface by conveying signals (transcription factors), that are then processed into output signals (mRNAs). WSNs operate in a similar manner, where sensor nodes send signals to others in the form of data packets. Packets at destination nodes convey forwarding instructions, which in return relays such packets to other sensors.

Recently, we have shown that wireless sensor networks adopting the transcriptional regulatory topologies (of E. coli), designated as bio-inspired WSNs, are more efficient than those adopting ER topologies of the same size in terms of conveying packets to sink nodes [28–30]. A support vector machine model was constructed with ~90% accuracy to predict packet receipt rates using the topological features of the networks as input [31,32] that includes the average degree, network density, as well as the abundance of FFLs. Each of these three topological features was ranked higher than the other ones. It is hence important to study how FFL abundance positively or negatively correlates with the other topological features in the network. Such a study will motivate the design of smart WSN topologies that exhibit similar FFL abundance and possibly FFL distribution as observed in the TRNs of E. coli and hence will have better efficiency in terms of their average packet receipt rates under node/link failures and channel noise.

2.2. Biological network growing algorithms

Another popular area of research includes the transcriptional network growing algorithms primarily based on the preferential attachment model [33]. Currently, only the TRNs of E. coli and yeast have been validated experimentally; hence such network growing algorithms are essential to allow the community to study the properties of such TRNs, design robust networks [34], as well as to predict the TRNs of higher-level organisms. We have recently developed such a network growing algorithm by extending the preferential attachment model to produce directed networks that mimic the topological properties of E. coli [35] in terms of their degree distribution and FFL abundance. The algorithm adds one foreign node at a time to an existing subgraph of candidate nodes. Based on in- and out-degree centralities of the candidate nodes, attachment kernels are computed [36], which further become conditions for foreign nodes attachments. Although this algorithm showed good correspondence in terms of FFL abundance in the predicted networks when compared to the TRN of E. coli, the corresponding distribution of FFLs measured in terms of the number of nodes in the network participating in such FFLs did not match well. Hence, while preferential attachment alone can only produce comparable degree distribution or motif abundance, additional topological metrics must be considered to design more accurate algorithms for growing TRNs. Our work in this paper, is hence important in this regard as we identify the topological metrics that strongly correlate with FFLs as observed from successive FFL deletions from the original TRN in E. coli.

3. The E. coli transcriptional network

We consider the TRN of E. coli, wherein $\gamma = 2.1 \pm 0.3$ [37]. The TRN was rendered using GeneNetWeaver [38]—a bioinformatics software tool which was originally designed to assess the accuracy of reverse engineered GRNs of E. coli and Yeast. E. coli is composed of 1565 nodes and 3758 interconnections, together forming 23 disjoint components.

A TRN is represented using a square adjacency matrix T. A direct connection originating from node i incident on node j in a TRN is represented by cell $T_{ij} = 1$, and the absence of such connection designates $T_{ij} = 0$. Because the edges in the TRN carry no weight, T can only hold values of 0 and 1.

A randomized instance of E. coli should have exactly the same number of nodes and edges, provided each node preserves its in- and out-degrees. The number of FFL substructures, N_F, in the real network is much higher than that of the randomized network, N_F^{rand}. The following conditions determine N_F's significance:

1. $N_F \geq 4$ for four or more disjoint instances of the FFL.
2. The probability for N_F^{rand} being greater than N_F is less than or equal to 1%.
3. For a large number of generated randomized instances of E. coli, $N_F - \langle N_F^{rand} \rangle > 0.1 N_F^{rand}$.

In a TRN of size n, we consider two methods for counting the occurrences of FFLs. The first method considers all 3 node combinations which form only FFLs and no other 3 node substructure; we designate such FFLs as the "canonical" FFLs. This can be computed using,

$$m = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \left[T_{ij} \cap T_{ik} \cap T_{jk} \right], \quad T_{ij} = T_{ik} = T_{jk} = 1,\ T_{ji} = T_{ki} = T_{kj} = 0. \tag{1}$$

The second method considers all available 3 node combinations forming at least one FFL; we designate such FFLs as the "embedded" FFLs. This can be determined computationally using,

$$m^* = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \left[T_{ij} \cap T_{ik} \cap T_{jk} \right], \quad T_{ij} = T_{ik} = T_{jk} = 1. \tag{2}$$

Fig. 1 illustrates the two types of FFLs considered in this paper. From our previous work, we have observed that m and m^* exhibit similar trends in correlation for all the topological features considered [39]. Hence, in this paper, we restrict the analysis to only canonical FFLs i.e., the m counts.

4. In-silico FFL knock-out algorithm

The schematic of the in-silico FFL knock-out algorithm is shown in Fig. 2. For simplicity, we designate the initial value of m as m_0 (i.e., at iteration zero) and m_j denotes the number of canonical FFLs at the jth iteration. The algorithm starts by computing m_0 and the different topological metrics discussed in the next section. Next, the network goes through a predefined number of iterations, wherein at each iteration we make a switch between two FFL edges (denoted in red in Fig. 2) that involve four distinct FFL nodes following the method proposed in [40]. The switch picks two edges at random from two different canonical FFLs, having two random source and destination nodes, s_1, s_2, d_1, and d_2 respectively. Edges are next rewired so that s_1 connects d_2 and s_2 connects d_1.

Every iteration is considered successful if the following two conditions are satisfied:

1. The rewiring of edges should not affect any other existing edge in the network. Doing so, the overall in-, out- and cumulative degree distributions stay exactly the same while the actual TRN topology is altered. Hence, the number of FFLs are changed.
2. For the $j+1$th iteration, $m_j < m_{j+1}$ must hold in order to proceed to the next iteration. This guarantees that one or more FFLs were deleted before starting the next iteration.

If either of the conditions are not satisfied, the iteration is repeated. This process continues for 200 iterations that was arbitrarily set however was enough to properly identify the correlations and trends that we report here. This process was repeated for 50 different sequence runs to remove the bias of any arbitrary FFL knock-out sequence on the reported correlations.

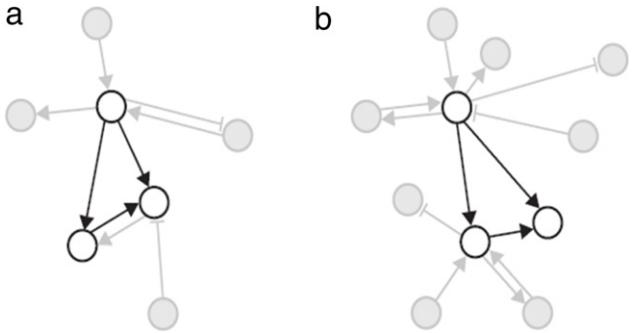

Fig. 1. Highlighted in bold: (a) Embedded and (b) Canonical FFLs; arrows denote up-regulation while blunt-ends denote down-regulation of nodes.

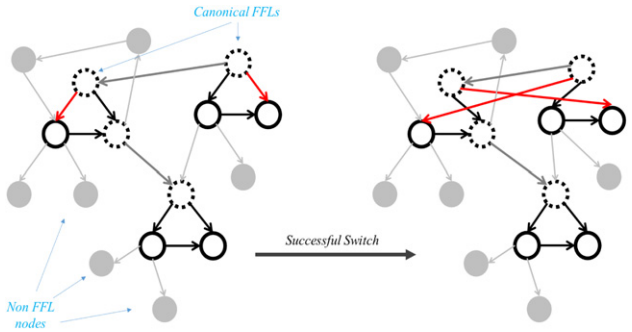

Fig. 2. Schematic of the in-silico FFL knock-out algorithm; red edges were switched to delete to canonical FFLs in this example. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Due to the randomness of the switches, each sequence will possibly exhibit a different trend in the FFL knock-out numbers at the successive iterations. This is because at each iteration, for a different FFL deletion sequence, 2 or more FFLs may be deleted depending on the level of interconnectedness of the FFLs; note that such canonical FFLs may share vertices with other canonical FFLs forming a set and deleting one edge of such an FFL can potentially destroy the whole set of canonical FFLs. In such cases we extrapolated the topological metric values such that mean of all 50 FFL deletion iterations could be reported on a common axis.

5. Topological metrics exhibiting strong correlation

We first report the four topological metrics that strongly correlate with the FFL knockouts. Other metrics were either not affected by the switches or showed slight correlations towards information transport or edge rewiring capabilities of the E. coli TRN and are discussed in the next section. The actual correlation matrices for all metrics are however discussed later in the paper. All figures plot the mean values of the topological metrics with corresponding error bars from 50 different runs of the FFL knock-out experiments.

5.1. Average shortest path

The average shortest path measures the average number of hops along the shortest paths between all possible node pairs in the network and is considered as one of the most important metrics for assessing network efficiency and robustness in terms of information transport [41]. The average shortest path s, is computed using,

$$s = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}}{n(n-1)}, \quad q_{ij} \neq \infty, \tag{3}$$

where q_{ij} is the shortest path between nodes i and j. $q_{ij} = \infty$ signifies that a path does not exist. Fig. 3 plots the successive FFL deletions vs. s and shows that s is negatively correlated with canonical FFLs. FFL structures decrease the average number of hops between the nodes in the TRN, and hence, play a major role in its efficiency.

5.2. Diameter

Although the maximum shortest path d is known to correlate positively with s, and hence negatively with FFL counts, we find d useful in other means. Since the metric considers the maximum of the paths, it should be more stable to changes in m. The points at which d increases are focal points for attention because they can help us pinpoint the more important FFLs. Fig. 4 shows E. coli's diameter, $s_0 = 6$.

5.3. Closeness centrality

The closeness centrality C_i measures the relative closeness of node i to every other node in the network. A close node is capable of delivering information quicker

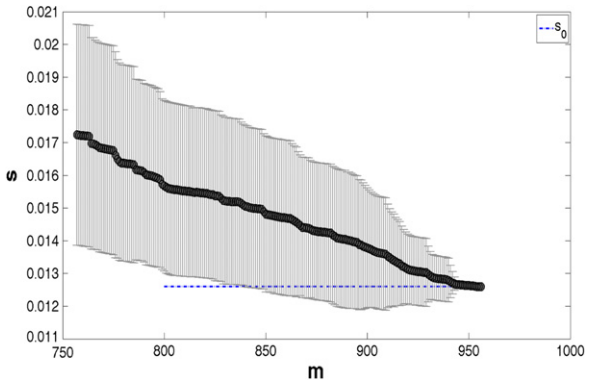

Fig. 3. The average shortest path, s, initially at s_0 goes through successive increments as the number of FFLs, m, decrease; the blue line designates s_0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

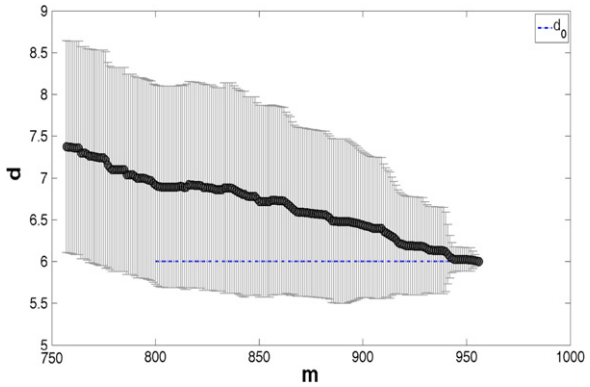

Fig. 4. The diameter, d, initially at d_0 goes through successive increments as the number of FFLs, m, decrease; blue line designates d_0. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to other nodes [42]. Central nodes in biological networks are crucial because they play the roles of 'organizational hubs' [43]. The closeness centrality of node i, C_i, is computed by,

$$C_i = \frac{n-1}{\sum_{i=1}^{n} q_{ij}}, \quad q_{ij} \neq \infty, \tag{4}$$

having values [0, 1], with 0 meaning i is disconnected from the network, and 1 meaning node i has a direct link to every node.

It is important to note that an increase in C_i signifies an increase in communicative efficiency, however, it does not guarantee resilience to random attacks. For example, if all nodes have $C_i = 1$, meaning each node is connected to one another by a direct link, a random attack on any node would cost the network $n-1$ links. For the entire network we compute the average closeness as,

$$\langle C \rangle = \frac{n-1}{n\sum_{i=1}^{n} q_{ij}}, \quad q_{ij} \neq \infty. \tag{5}$$

Fig. 5 depicts a negative correlation between the average closeness centrality $\langle C \rangle$ and the canonical FFL abundances, m. FFLs seem to preserve the 'scale-free' property of the network wherein a decrease in FFL counts cause an increase in the average closeness, which further means that the hub node strengths shift to other nodes in the network. The availability of centralized nodes fades as the FFLs are deleted, making the network more communicatively efficient, however, less resilient to random attacks.

5.4. Average local clustering coefficient

The average local clustering coefficient [10], $\langle L \rangle$, formally defined as

$$\langle L \rangle = \frac{1}{n}\sum_{i=1}^{n} \frac{\sum_{j=0}^{n} T_{ij}}{k_i(k_i-1)}. \tag{6}$$

k_i denotes the number of neighbors for node i, while the metric measures the number of edges connecting the nodes as a fraction of the number of possible edges that could exist in their local communities. At the individual node level, this metric can be used to quantify relative nodal participation in embedded clusters. Fig. 6 shows a positive correlation with canonical FFLs, meaning there is a decrease in the number of local communities as FFLs get deleted. This supports the observation stated by [44,45] that scale-free networks typically tend to form clusters.

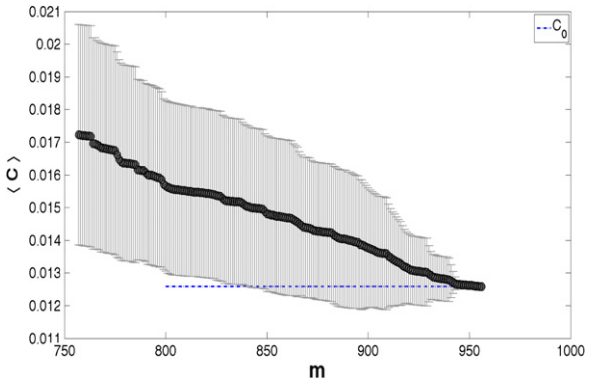

Fig. 5. Changes in the average closeness centrality, $\langle C \rangle$, of E. coli's TRN as a function of the abundance of canonical FFLs, m.

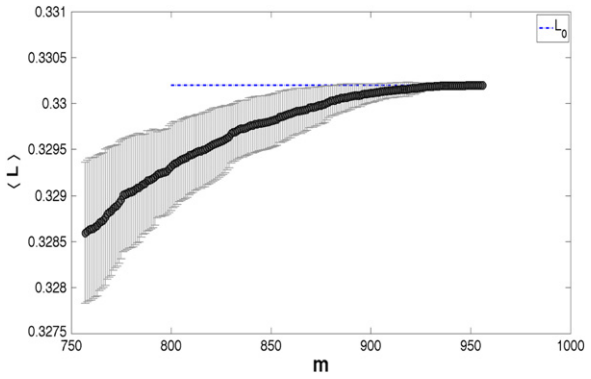

Fig. 6. Changes in the average local clustering coefficient, $\langle L \rangle$, of E. coli's TRN as a function of the abundance of canonical FFLs, m.

6. Metrics exhibiting slight or no correlation

In this section we report the topological metrics exhibiting little or no correlation to canonical FFLs. We will first demonstrate a slight correlation between m and a fifth connectivity related metric, namely the global clustering coefficient. In addition, we establish that there is no direct relation between FFL abundance and nodal resilience or end-to-end information transport delays.

6.1. Global clustering coefficient

Clustering coefficient measures the tendency for a network to form clusters. Evidence shows that real-world networks create denser ties than ER networks [10]. Moreover, robust biological networks having scale-free distributions, exhibit short average paths and high clustering coefficients [44,45]. We denote the global clustering coefficient by G, which is computed by dividing thrice the number of completely connected nodes by the number of triplets (i.e. open loops).

Fig. 7 shows the changes in G by our FFL knockout experiments. From G_0, there is a noticeable oscillating decline in the curve towards lower m values, however the correlation is not strong. Moreover, the ranges of the oscillations in terms of the standard deviations increase as m increases. These observations suggest that this metric will need separate consideration in designing better TRN growing algorithms. In general, such TRN growing algorithms try to preserve the degree and motif distributions and need to additionally consider the global clustering coefficients of the generated networks to properly mimic the topological characteristics of real TRNs of organisms.

6.2. Node failures

A random node removal experiment can assess fault tolerance robustness [6] in networks. Typically after a number of successive node failures, the network is subject to being disconnected; each node failure typically weakens the network's connectivity by eliminating paths. For example, the Internet is known to have high fault tolerance because it can resist multiple failures yet relay packets to destination terminals [7]. In an experiment comparing scale-free to exponential networks, the size the of the diameter for scale-free networks barely increased as compared to that of the exponential [6]. This experiment demonstrates that scale-free networks are more robust to node failures.

In Fig. 8, two networks of the same number of edges and nodes are observed for the size of their largest connected components (LCC) post random nodal failures. The TRN of E. coli and a randomly generated network were considered for this experiment. The random network was generated using an iterative process of selection and connection:

- **Selection:** pick a random source and destination node that have not been selected before.

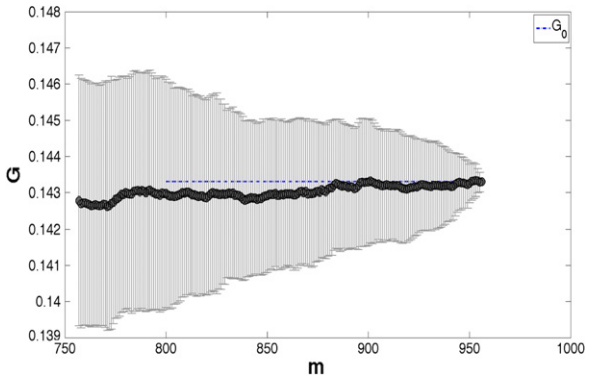

Fig. 7. Changes in the global clustering coefficient, G, of E. coli's TRN as a function of the abundance of canonical FFLs, m.

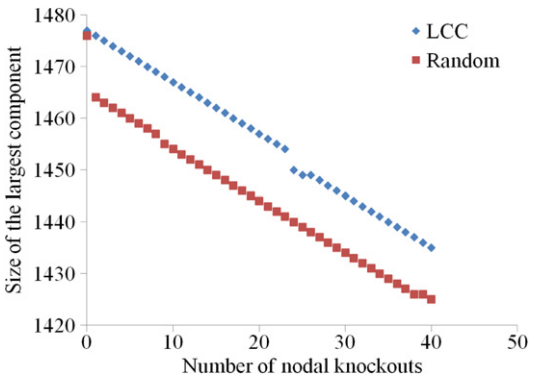

Fig. 8. An experimental plot of E. coli's LCC and a randomly generated network of the same size. Each point represents the current size of their largest connected component as a function of the number of random node deletions.

- **Connection:** draw a directed edge from source to destination.

This process is repeated for the same number of edges as that in the LCC of the TRN of E. coli. A linear best fit shows that the random network degrades at a higher rate thereby suggesting the higher robustness to node failures for E. coli's TRN. This can be due to the fact that LCC is sparse, having 1477 nodes and 3671 edges making any random network generated using these constraints be sparse as well.

The next question we try to answer is whether the canonical FFLs impart such robustness to node failures in E. coli's TRN. For each of the 50 FFL knock-out sequences, the gradient of the linear best fit, K, of the largest connected component was recorded post FFL knockouts. At each FFL knock-out step, 40 random node deletions were then simulated to estimate K; these node deletions were then rolled back before the next FFL deletion step. This experiment demonstrates a weak negative correlation between K and m (Fig. 10; cell (m, K)) that supports the case with $\langle C \rangle$, where it was observed that abundance of FFLs gave higher possibility to the occurrence of hub nodes—a condition which should ultimately decrease K. This demonstrates that canonical FFLs may not directly contribute to fault tolerance of the network based on node failures. However,

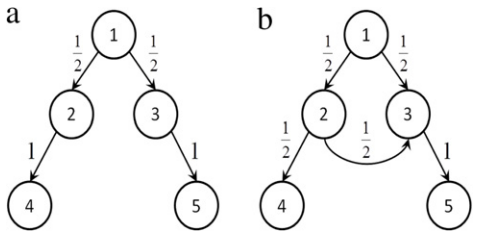

Fig. 9. Schematic of the mRNA first passage time dynamics; mRNAs start from node 1 and reach the leaf nodes (4, 5) after successive hops. Values on each edge designate the probability for the mRNA to traverse the corresponding link.

definitive conclusions regarding K cannot be drawn from such FFL deletion experiments alone without exploring other metrics such as the diameter, and other approaches such as edge knockouts. Note that canonical FFLs only provide fault-tolerance for the intermediate node knock-out due to the direct edge from the top to the third node; hence, the probability of selecting this intermediate node in the nodal knock-out experiments was quite low resulting in the weak negative correlation that we observed.

6.3. Single mRNA expected first passage time

The final metric we considered is based on a simplified random walk model for the dynamics of information transport in a TRN. This model for a single mRNA to traverse a TRN considers three types of nodes:

1. **Transmitter** nodes having no out-degree.
2. **Intermediate** nodes having both out- and in-degrees.
3. **Receivers** nodes having no in-degree.

Moreover, the half-life of mRNA degradation is not considered, i.e. in every case an mRNA initiated from a transmitter will eventually reach the receiver. The sample TRN's shown in Fig. 9 show the effects of a canonical FFL on the probability calculations of the links for a traversing mRNA. The sum of the probabilities on all outgoing edges from a node equals 1 to ensure normalization. mRNAs use a random walk to traverse the TRN until they reach the leaf nodes.

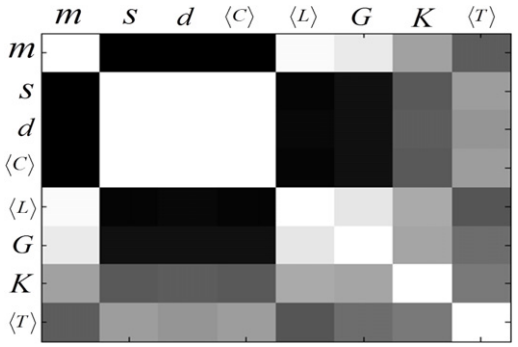

Fig. 10. Gray-scale Pearson correlation matrix between the different metrics.

A computational experiment was designed to compute the mRNA Expected First Passage Time, T in a TRN. The edge probabilities were generated following a uniform distribution based on the number of outgoing links from each node. A random transmitter node is chosen initially that will generate the mRNA. The mRNA takes a random walk towards the nearest neighbors based on the edge probabilities and the process continues till the mRNA reaches a receiver. This process is repeated for 10 000 different instances (each with a randomly chosen initial transmitter node) and the mean path length (in number of hops) is recorded in $\langle T \rangle$. This whole process was repeated at each canonical FFL knock-out step.

Ideally, the mean first passage time of mRNAs should be directly correlated to the average shortest path in the network. We did observe a weaker negative correlation between $\langle T \rangle$ and the abundance of canonical FFLs, m in comparison to the correlation values of average shortest path with m (Fig. 10; cell $(m, \langle T \rangle)$). This experiment conceptually simulates a flooding type routing protocol in TRNs however, as each mRNA is transmitted in a separate run, the effects of congestion at the nodes/links and mRNA degradation were not captured. In spite of this drawback, this experiment validates our observations on the strong negative correlation between the average shortest path and canonical FFL abundance thereby underlining their importance in efficient information transport in TRNs.

7. Correlation between different metrics

As our primary goal is to study the impact of canonical FFLs on the different robustness metrics in TRNs, we reported the correlation of FFL abundance to each of these metrics above. However, our FFL knock-out experiments also allow us to compute the pair-wise correlation between these eight metrics and FFL abundance using a 8×8 correlation matrix. This correlation matrix was computed using Pearson's correlation coefficient where the values range from -1 to 1, with 1 showing perfect positive correlation, -1 showing perfect negative dependency, and 0 meaning completely independent of one another. A Matlab module [46], colormap(gray) was then used to illustrate the gray scale matrix of the correlation coefficients in Fig. 10. Colors closer to black correspond to almost -1 correlation, and white for 1.

The figure summarizes the impact of the canonical FFLs on the other 7 metrics (first column in the correlation matrix). We observe a strong positive correlation with the average local clustering coefficient and strong negative correlation with average shortest path, diameter and average closeness centrality metrics. While our modified preferential attachment model for TRN growth conceptually guarantees high local clustering coefficients and motif abundance, it does not inherently impact the average shortest path or average closeness centrality. So, should one consider both of these metrics in addition to preferential attachment to design new TRN growth algorithms? The correlation matrix however shows a strong positive correlation between average shortest path and closeness centralities suggesting that consideration of any one of these two metrics will be sufficient as long as the TRN growth algorithm generates the right degree and motif distributions.

The second column suggests that the average shortest path negatively correlates with the average local clustering coefficient. This shows that a preferential attachment should be biased towards reducing number of hops between pairs of transcription factors and genes while enhancing local community formations. Previous works have addressed robustness for networks due to their ability to preserve fault tolerance due to their abundance in local modules [41], hence both metrics come hand in hand. Average closeness centrality shows similar correlations with local clustering, which also makes it a candidate for preferential attachment. In all cases, growth models should be examined separately for attachment kernels combining both $\langle L \rangle$ and s, as well as $\langle L \rangle$ and $\langle C \rangle$. Note that the diameter is directly correlated to average shortest path and need not be considered separately.

The global clustering coefficient slightly correlates with the mentioned metrics (columns 1–5), with correlations exceeding 80% in each case. This is not as strong as the first 5 columns having correlations exceeding 95% between each of their pairs. This suggests that the global clustering must be checked separately when adding each node to the substrate network using the preferential attachment. Because $\langle G \rangle$ positively correlates with m, only node additions resulting in an increase of $\langle G \rangle$ should be allowed.

Other metrics such as single mRNA mean passage time and the gradient of the knockout curves show weak correlations with rest, having correlations less than 32%. However, the slight correlations with m in columns 7 and 8 suggest that canonical FFL abundance assists in robustness in two ways: fault-tolerance to node failures and information transport efficiency based on first passage times. A slight positive correlation between m and K shows that FFL abundance only slightly aids in keeping the network intact post random attacks. Moreover, a slight negative correlation between m and T explains how lower average shortest paths contributed by canonical FFLs may aid in information transport robustness.

With regard to bettering the transmission efficiency of bio-inspired wireless sensor networks, enhancements can be achieved by considering both correlating and non-correlating metrics. For example, a support vector machine (SVM) model that considers input features

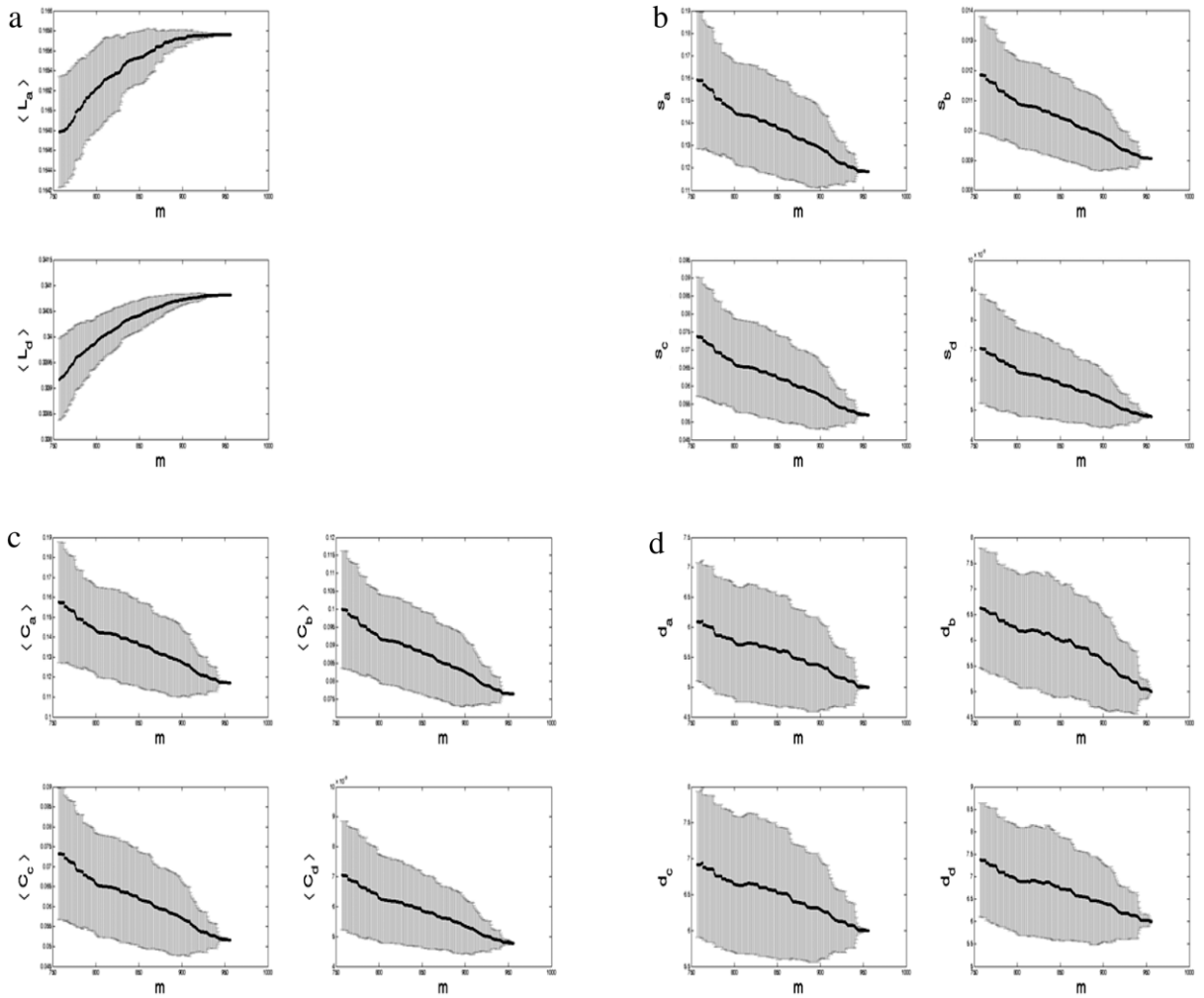

Fig. 11. Changes in the sub-metrics of the: (a) local clustering coefficient $\langle L \rangle$, (b) average shortest path (s), (c) average closeness centrality $\langle C \rangle$ and (d) diameter (d), against the changes in canonical FFL abundance, m.

derived from topological measures coupled with average packet receipt percentages, can be trained to predict network transmission efficiency based on such topological features [31]. Such models can be created faster by reducing the input feature space, which can be achieved by only considering strongly correlated metrics like those of columns 1–5. For example, features derived from $\langle C \rangle$ should cover for the effects of s, D, and L in training the model. This would decrease the SVM prediction plane by 3 dimensions making it less complex and more accurate. On the other hand, the results reported here will help one to realize other features that can be added to this feature space because they are weakly correlated with the other features.

8. Zooming into the strongly correlating metrics

A deeper understanding of how the strongly correlating metrics are related to m would require us to identify some sub-metrics by dividing the TRN into four categories. Each of the path-based metrics s, d, and $\langle C \rangle$ can be divided into four different categories based on the type of source–destination node pairs. In terms of terminology, the metric X can be designated in the following ways:

1. X_a refers to the metric X computed exclusively using the intermediate nodes of the network.
2. X_b refers to the metric X computed exclusively using the intermediate nodes as source and receiver nodes as destination in the network.
3. X_c refers to the metric X computed exclusively using the transmitter nodes as source and intermediate nodes as destination in the network.
4. X_d refers to the metric X computed exclusively using the transmitter nodes as source and receiver nodes as destination in the network.

The metric $\langle L \rangle$ can however be sub-divided into two categories based on whether the average local clustering coefficient is exclusively computed for (i) the intermediate nodes and (ii) the peripheral (transmitter or receiver) nodes (see Fig. 11).

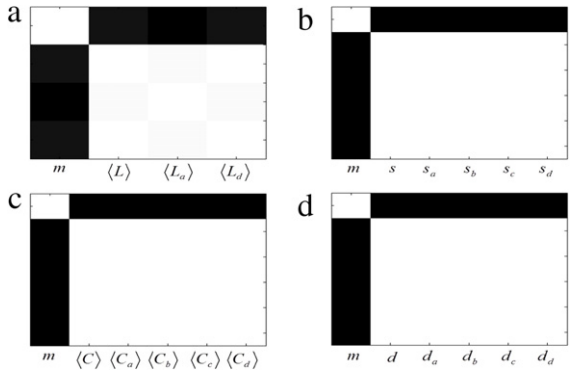

Fig. 12. Gray-scale correlation matrix for the sub-metrics: X_a, X_b, X_c, and X_d. Metric X references (a) $\langle L \rangle$, (b) s, (c) $\langle C \rangle$, and (d) d respectively. The second column in each panel references the actual metric X.

Fig. 12 shows how each of the sub-metrics correlates with canonical FFL abundance m using our in-silico FFL knock-out experiments. Each of the sub-metrics belonging to a particular topological feature shows exactly the same trend in terms of positive or negative correlation and hence is hard to differentiate based on the plot. Fig. 12 shows the actual correlation matrix in gray-scale for each of the sub-metrics; it corroborates the fact the each sub-metric shows very similar positive or negative correlations. These observations point to the fact that canonical FFLs are *evenly* distributed in the TRN possibly involving similar percentages of peripheral or intermediate nodes; hence the impact of random FFL deletions will similarly impact the shortest paths, diameters or closeness centralities of peripheral, intermediate or mixed node-pairs. The local clustering coefficients are based on individual nodes and are also similarly impacted for peripheral or intermediate nodes.

To confirm this observation, we next enumerated the shortest paths for each sub-metric belonging to s from the TRN in E. coli. Fig. 13 shows the percentage of edges from the shortest paths that participate in canonical FFLs. It seems that the canonical FFLs on the periphery of the TRN (green line denoting transmitter–intermediate node pairs) as well as the intermediate FFLs (red line denoting intermediate–intermediate node pairs) have the highest contributions towards the over-all shortest path in the TRN. While this still suggests that both peripheral and intermediate canonical FFLs are evenly distributed in the TRN, there is a possibility that the shortest paths are contributed by traversing the node-1 → node-3 direct edges of multiple canonical FFLs embedded in the TRN.

Fig. 14 shows the percentage of direct neighbors of each node that participate in canonical FFLs in the TRN of E. coli. There is not much separation in these plots suggesting that both peripheral and intermediate canonical FFLs contribute equally to the average local clustering coefficient of the entire TRN.

9. Conclusions

In this paper, we designed an innovative feed-forward loop motif knockout experiment to assess the impact of

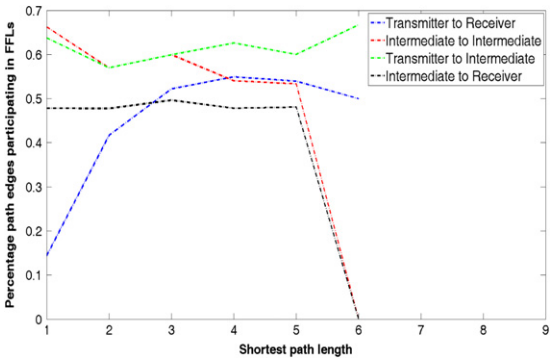

Fig. 13. Percentage of edges participating in canonical FFLs in the shortest paths in the TRN of E. coli.

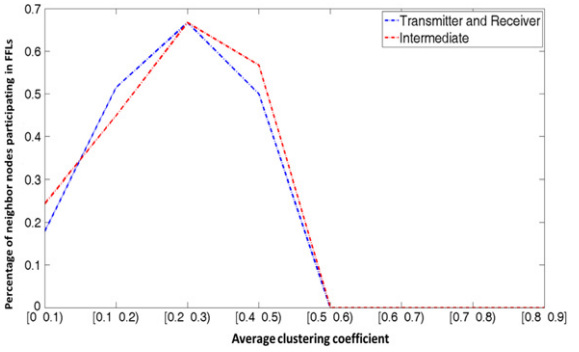

Fig. 14. Percentage of neighbors participating in canonical FFLs for nodes having local clustering coefficients within ranges shown in the horizontal axis for the TRN in E. coli.

such FFL deletions on other topological metrics in the TRN of E. coli. The purpose of this study is two-fold: motivate the design of more accurate TRN growing algorithms as well as design more efficient bio-inspired wireless sensor network topologies. As similar characteristics were observed for embedded and canonical FFLs, we only considered in-silico canonical FFL deletions in this paper. FFL knockouts show strong negative correlation with average shortest path and average closeness centrality. As these two metrics are related and contribute to lower end to end delays for information transport, such negative correlation indirectly illustrates the importance of FFLs in facilitating the information transport in such networks. On the other hand FFL knockouts show a strong positive correlation with the average local clustering coefficients in the network; this suggests that canonical FFLs play an important role in maintaining the high clustering coefficients and hence in the formation of local communities in the network.

We further observed a weak correlation between canonical FFLs and fault-tolerance in such networks measured in terms of random node deletions. This suggests that canonical FFLs may not contribute as much in maintaining such fault-tolerance in TRNs. Similar weak correlation was observed for mean first passage transport times of mRNAs in a TRN; hence canonical FFLs may facilitate shortest path based routing instead of flooding in TRNs.

We further categorized the shortest path, closeness centrality, diameter and local clustering coefficient metrics into several sub-metrics based on peripheral or intermediate nodes. It appears that canonical FFLs are evenly distributed in the TRN as their contributions to these metrics are independent of the whether the FFLs themselves are located at peripheral or intermediate positions in the TRN. However, multiple canonical FFLs seem to participate in the source–destination shortest paths where routing is facilitated by stitching together the direct edges of these FFLs.

Acknowledgments

This research was funded by the US Army's Environmental Quality and Installations 6.1 basic research program, under contract W912HZ-14-P-0008. Opinions, interpretations, conclusions, and recommendations are those of the author(s) and are not necessarily endorsed by the U.S. Army.

References

[1] H. Kitano, Biological robustness, Nat. Rev. Genet. (2004) 826–837.
[2] H. Kitano, Towards a theory of biological robustness, Mol. Syst. Biol. 3 (137) (2007).
[3] I.F. Akyildiz, I.H. Kasimoglu, Wireless sensor and actor networks: research challenges, Ad Hoc Networks 2 (4) (2004) 351–367.
[4] C. Genio, T. Gross, K.E. Bassler, All scale-free networks are sparse, Phys. Rev. Lett. 107 (2011).
[5] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (4) (2009) 661–703.
[6] R. Albert, H. Jeong, A.-L. Barabási, Error and attack tolerance of complex networks, Nature 406 (2000) 378–382.
[7] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Resilience of the Internet to random breakdowns, Phys. Rev. Lett. 85 (2000) 4626–4628.
[8] R. Cohen, K. Erez, D. ben Avraham, S. Havlin, Breakdown of the Internet under intentional attack, Phys. Rev. Lett. 86 (2001) 3682–3685.
[9] P. Erdös, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960).
[10] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[11] S. Magnan, U. Alon, Structure and function of the feed-forward loop network motif, Proc. Natl. Acad. Sci. USA (2003).
[12] J.M. Rip, K.S. McCann, D.H. Lynn, S. Fawcett, An experimental test of a fundamental food web motif, Proc. Biol. Sci. (2010).
[13] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: Simple building blocks of complex networks, Science 298 (5594) (2002) 824–827. 25.
[14] J.J. Faith, B. Hayete, J.T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J.J. Collins, T.S. Gardner, Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles, PLoS Biol. 5 (1) (2007) e8.
[15] A.A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R.D. Favera, A. Califano, ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, BMC Bioinformatics 7 (Suppl 1) (2006) S7.
[16] I. Shmulevich, E.R. Dougherty, Probabilistic Boolean Networks—The Modeling and Control of Gene Regulatory Networks, SIAM, 2010.
[17] J. Feng, J. Jost, M. Qian, Networks: from Biology to Theory, Springer, 2007.
[18] R.J. Prill, P.A. Iglesias, A. Levchenko, Dynamic properties of network motifs contribute to biological network organization, PLoS Biol. (2005).
[19] S.S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of Escherichia coli, Nature Genet. 31 (2002) 1061–4036.
[20] T. Feyessa, M. Bikdash, Measuring nodal contribution to global network robustness, in: Southeastcon, 2011 Proceedings of IEEE, March 2011, pp. 131–135.
[21] P. Crucittia, V. Latorab, M. Marchioric, A. Rapisardab, Error and attack tolerance of complex networks, Physica A 340 (2004) 388–394.
[22] V. Latora, M. Marchiori, Efficient behavior of small-world networks, Phys. Rev. Lett. 87 (2001) 198701.
[23] J. Efstathiou, A.K. Ng, Structural robustness of complex networks, in: Proceedings of International Workshop and Conference on Network Science, NetSci 2006, May 2006.
[24] V. Agoston, P. Csermely, S. Pongor, Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example, Phys. Rev. E (2005).
[25] J.-R. Kim, Y. Yoon, K.-H. Cho, Coupled feedback loops form dynamic motifs of cellular networks, Biophys. J. 94 (2008) 359–365.
[26] Y.Keun Kwon, K.Hyun Cho, Boolean dynamics of biological networks with multiple coupled feedback loops, 2007.
[27] M. Isalan, C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao, E. Raineri, M. Garriga-Canut, L. Serrano, Evolvability and hierarchy in rewired bacterial gene networks, Nature 452 (7189) (2008) 840–845.
[28] B. Kamapantula, A. Abdelzaher, P. Ghosh, M. Mayo, E. Perkins, S. Das, Leveraging the robustness of genetic networks: a case study on bio-inspired wireless sensor network topologies, J. Ambient Intell. Humaniz. Comput. 5 (3) (2014) 323–339.
[29] B. Kamapantula, A. Abdelzaher, P. Ghosh, M. Mayo, E. Perkins, S. Das, Performance of wireless sensor topologies inspired by e. coli genetic networks, in: PerCom Workshops, 2012, pp. 302–307.
[30] P. Ghosh, M. Mayo, V. Chaitankar, T. Habib, E. Perkins, S. Das, Principles of genomic robustness inspire fault-tolerant WSN topologies: A network science based case study, in: PerCom Workshops, 2011, pp. 160–165.
[31] A. Abdelzaher, B.K. Kamapantula, P. Ghosh, S.K. Das, Empirical prediction of packet transmission efficiency in bio-inspired wireless sensor networks, in: A. Abraham, A.Y. Zomaya, S. Ventura, R. Yager, V. Snásel, A.K. Muda, P. Samuel (Eds.), ISDA, IEEE, 2012, pp. 705–710.
[32] B. Kamapantula, A. Abdelzaher, M. Mayo, E. Perkins, S. Das, P. Ghosh, Quantifying robustness in biological networks using ns-2, Springer-Nanocom, 2014.
[33] A.-L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[34] S. Ghosh, P. Ghosh, K. Basu, S.K. Das, GaMa: An evolutionary algorithmic approach for the design of mesh-based radio access networks, in: 30th Annual IEEE Conference on Local Computer Networks (LCN 2005), IEEE, 2005, pp. 374–381.
[35] M. Mayo, A.F. Abdelzaher, E.J. Perkins, P. Ghosh, Motif participation by genes in e. coli transcriptional networks, Front. Physiol. 3 (2012) 357.
[36] P.L. Krapivsky, S. Redner, F. Leyvraz, Connectivity of growing random networks, Phys. Rev. Lett. 85 (21) (2000) 4629–4632.
[37] A. Vazquez, R. Dobrin, D. Sergi, J.P. Eckmann, Z.N. Oltvai, A.L. Barabasi, The topological relationship between the large-scale attributes and local interaction patterns of complex networks, Proc. Natl. Acad. Sci. USA 101 (52) (2004) 17940–5+.
[38] T. Schaffter, D. Marbach, D. Floreano, GeneNETWEAVER: in silico benchmark generation and performance profiling of network inference methods, Bioinformatics 27 (5) (2011) 2263–2270. Aug 15.
[39] A. Abdelzaher, M. Mayo, E.J. Perkins, P. Ghosh, Correlating in-silico feed-forward loop knockout experiments with the topological features of transcriptional regulatory networks, in: International Conference on Bio-Inspired Information and Communications Technologies, November 2014.
[40] M.E.J. Newman, S.H. Strogatz, D.J. Watts, Random graphs with arbitrary degree distributions and their applications, Phys. Rev. E 64 (2001) 026118.
[41] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.
[42] M.E.J. Newman, A measure of betweenness centrality based on random walks, Social Networks (2005).
[43] S. Wuchty, P.F. Stadler, Centers of complex networks, J. Theor. Biol. 7 (2003) 45–53.
[44] K.-I. Goh, B. Kahng, D. Kim, Graph theoretic analysis of protein interaction networks of eukaryotes, Physica A 357 (3) (2005) 501–512.
[45] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.L. Barabási, Hierarchical organization of modularity in metabolic networks, Science 297 (5586) (2002) 1551–1555.

[46] MATLAB. version 7.10.0 (R2010a). The MathWorks Inc., Natick, Massachusetts, 2010.

Ahmed F. Abdelzaher (abdelzaheraf@vcu.edu) is a doctoral student at the Department of Computer Science, Virginia Commonwealth University. Prior to this, he completed his B.S and M.S. studies in Computer Science from the School of Computing, University of Southern Mississippi. His research interests include data mining, algorithm development, modeling and simulation in the biological networks and bio-inspired wireless networking domain areas.

Michael L. Mayo (Michael.L.Mayo@usace.army.mil) is a Research Physicist at the Environmental Lab, US Army Engineering Research and Development Center, Vicksburg, MS. He completed his B.S., M.S. and Ph.D. degrees in Physics from the University of Missouri at Columbia. His research interests include mathematical modeling of biological systems, PK type modeling, stochastic (SSA algorithms), statistical physics modeling (diffusion–reaction systems).

Edward J. Perkins (Edward.J.Perkins@usace.army.mil) is a Senior Research Scientist at the Environmental Lab, US Army Engineering Research and Development Center, Vicksburg, MS. He completed his B.S. in Genetics from University of Illinois at Champaign and Ph.D. in Genetics and Cell Biology from Washington State University, Pullman. His research interests include environmental toxicology, habitat creation and rehabilitation, decision analysis and risk assessment, invasive species, threatened and endangered species, molecular diagnostics, environmental biotechnology, impact of military stressors on individuals and populations and identification of toxicants and impacts using gene expression markers.

Preetam Ghosh (pghosh@vcu.edu) is an Associate Professor at the Department of Computer Science, Virginia Commonwealth University. He received his B.S. from Jadavpur University, India and M.S./Ph.D. from the University of Texas at Arlington in Computer Science and Engineering. He directs the Biological Networks Lab at VCU and conducts research in stochastic modeling and simulation and optimization algorithms in various domains related to biomedical sciences and bio-inspired networking.